

# **Intrinsically Disordered Landscapes for Human CD4 Receptor Peptide**

Jerelle A. Joseph and David J. Wales\*

*Department of Chemistry, University of Cambridge, Lensfield Road,  
Cambridge CB2 1EW, United Kingdom*

E-mail: [dw34@cam.ac.uk](mailto:dw34@cam.ac.uk)

## Abstract

Due to their inherent structural plasticity, intrinsically disordered proteins (IDPs) are generally difficult to characterise, both experimentally and via simulations. In this work, an approach for studying IDPs within the potential energy landscape framework is implemented and tested. Specifically, human CD4 receptor peptide, a disordered region implicated in HIV-1 infection, is characterised via basin-hopping parallel tempering and discrete path sampling. We also investigate the effects of three state-of-the-art AMBER forcefields (ff99SB-ILDN, ff14ipq, and ff14SB) on the energy landscape. The results for ff99SB-ILDN exhibit the best agreement with experiment. Metastable states identified on the free energy surface help to unify, and are consistent with, several earlier predictions, and may serve as starting points for probing the reaction interface between CD4 and HIV-1 accessory proteins.

## Introduction

In the mid-1990's, the suggestion that many functional proteins were natively disordered was met with much scepticism.<sup>1,2</sup> This notion was in direct conflict with the linear sequence–structure–function paradigm that had guided molecular biology for over half a century. Since then, overwhelming evidence from experiment<sup>3–7</sup> and bioinformatics surveys of entire genome sequences<sup>8,9</sup> have driven a paradigm shift, and the abundance and importance of intrinsically disordered proteins (IDPs) or regions (IDRs) are now widely recognised. It has been estimated that about 50% of mammalian proteins contain contiguous disordered regions (>30 residues long) and about 25% are fully disordered.<sup>10</sup> Disordered sequences are generally characterised by a low content of bulky hydrophobic amino acids, a high proportion of polar and charged amino acids, and an overall low complexity (low complexity sequences contain multiple repeats of single amino acids or amino acid motifs).<sup>4,11,12</sup> Due to their composition, IDPs usually do not collapse to form hydrophobic cores, or fold into stable three-dimensional structures.<sup>1,7</sup> Instead, under physiological conditions, they display a high degree of conforma-

tional flexibility, in which several states rapidly interconvert.<sup>13,14</sup> Accordingly, IDPs exhibit diverse functional modes: in some cases, their functions depend directly on the disordered state,<sup>5,15</sup> while, in other instances, they undergo induced folding upon interaction with a molecular partner.<sup>3,16,17</sup>

These characteristics confer several functional advantages to IDPs, such as the ability to regulate distance and orientation of protein domains, bind to targets with high specificity and low affinity, interact with multiple targets, and undergo post-translational modifications. IDPs therefore play a central role in many fundamental processes, including transcription regulation, translation, cell signalling and phosphorylation.<sup>7,15,18,19</sup> On the other hand, intrinsic disorder in proteins “can also have a biological cost in terms of the promotion and proliferation of protein folding diseases”.<sup>7</sup> Specifically, the altered expression of IDPs has been linked to several diseases,<sup>20</sup> including cancer, diabetes, neurodegenerative diseases, and cardiovascular diseases.<sup>21–26</sup> Additionally, structural disorder has been associated with the action of pathogens;<sup>1</sup> for example, certain viruses may mimic IDRs and interfere with their regulation inside the cell.<sup>27</sup>

Thus, characterisation of IDPs is an important research area, to elucidate their functions and to identify potential therapeutic targets. Experimental characterisation of IDPs via conventional techniques has proved to be challenging, largely due to the fact that these methods were originally optimised for folded proteins. In particular, due to their dynamic nature, the electron density of IDRs is often absent from X-ray diffraction maps.<sup>5</sup> Additionally, in standard NMR studies, IDPs usually form aggregates at the required experimental concentrations, interconversion of conformational states leads to line-broadening effects, and peaks corresponding to disordered proteins are often poorly dispersed.<sup>28</sup> Techniques such as near and far-UV circular dichroism (CD) have been employed to distinguish between folded and disordered proteins. Since intrinsic disorder is commonly confined to a small region of the protein, CD methods, which lack residue-specific information, need to be combined with other techniques to properly characterise IDRs.<sup>5</sup> Mutidimensional NMR approaches,<sup>28,29</sup>

SAXS,<sup>30,31</sup> single-molecule FRET<sup>32</sup> and AFM<sup>33</sup> have all proved useful in providing structural and dynamical information for disordered proteins. In one study, multidimensional NMR was used in conjunction with isotopic labelling to probe IDPs *in vivo*.<sup>34</sup> Other hybrid approaches, such as protease digestion with mass spectrometry,<sup>35</sup> have also been employed to identify IDPs, albeit less effectively than NMR-based techniques.

Molecular simulations, which can provide high resolution structural and kinetic information for IDP ensembles, have complemented experimental studies.<sup>36-39</sup> In particular, these simulations can provide predictions for experiment, or can be used in conjunction with experiments to aid interpretation.<sup>40,41</sup> The quality of such predictions and interpretations depends primarily on the accuracy of protein force fields and on the efficiency of the sampling strategy employed. As for conventional experimental techniques, protein force fields were originally developed for folded globular proteins; thus, they may display secondary structure biases.<sup>42,43</sup> Hence, it is necessary to develop force fields that can achieve a good balance among secondary structures. To this end, several standard force fields have been modified and benchmarked for IDPs.<sup>44-51</sup> While most modifications have led to improved performance, several studies have yielded inconsistent results, and there is currently no general consensus on the best force field for IDPs.

Since the conformational space of disordered proteins is significantly more complex and heterogeneous than for globular proteins, enhanced sampling for IDPs is critical.<sup>52-54</sup> Accordingly, various enhanced sampling schemes and algorithms have been employed, such as temperature replica exchange molecular dynamics (tREMD),<sup>55</sup> bias-exchange metadynamics (BE-META),<sup>56-59</sup> and Markov state models (MSMs).<sup>60-65</sup> In replica exchange protocols the protein undergoes a random walk, where information is exchanged between replicas periodically.<sup>66</sup> Metadynamics based approaches may use experimental data to guide sampling,<sup>67</sup> whereas in MSMs the goal is to probe long time dynamics by constructing kinetic models from numerous (mainly short) unbiased MD simulations.<sup>68</sup> This formalism is based on a master equation approach, as in the present work. However, instead of molecular dynamics,



we characterise states and interconversion rates using geometry optimisation, with post-processing using tools of statistical mechanics and unimolecular rate theory. Several other sampling approaches have been used to characterise IDP ensembles with varying degrees of success,<sup>52–54</sup> including some that utilise reweighing techniques.<sup>41</sup>

In the present work, the cytoplasmic tail of human cluster of differentiation 4 (CD4), an IDR linked to HIV-1 infection, is investigated computationally.<sup>69</sup> CD4 is a glycoprotein (gp), mainly expressed on the surface of regulatory T cells (a subclass of white blood cells).<sup>70</sup> The glycoprotein contains 433 residues and is made up of three regions: an extracellular N-terminal region (371 residues), a transmembrane helix (24 residues) and a cytoplasmic C-terminal domain (38 residues), Figure 1.<sup>69,71,72</sup> The T cell receptors (TCR) are responsible for recognising antigen peptides bound to Major Histocompatibility Complex class II (MHC-II) molecules. During this recognition process, CD4 functions as a coreceptor, whereby the extracellular region binds to a region of MHC-II.<sup>73,74</sup> Additionally, the cytoplasmic tail of CD4 interacts with a lymphocyte-specific protein kinase (p56<sup>lck</sup>), by jointly coordinating Zn<sup>2+</sup> via a pair of cysteines on each molecule (residues 420 and 422 in CD4).<sup>75</sup> Ultimately, the interaction between CD4 and the TCR complex recruits p56<sup>lck</sup> to TCR (which induces phosphorylation of TCR-associated molecules),<sup>76,77</sup> and is a central upstream event in TCR signal transduction and the immune response.

CD4 also acts as a primary receptor of human immunodeficiency virus type 1 (HIV-1).<sup>78–81</sup> Specifically, entry of HIV-1 into the T cells is initiated by binding of the glycoprotein gp120 (on the HIV-1 envelope) to the extracellular region of CD4.<sup>82</sup> This initial interaction eventually leads to fusion of the viral and cell membranes and subsequent infection. Once the cell has become infected, the presence of CD4 is known to interfere with the viral life cycle;<sup>83,84</sup> for example, CD4 may inhibit the release of nascent viruses.<sup>85,86</sup> However, HIV-1 has evolved to produce viral proteins, Nef (negative factor) and Vpu (viral protein U), which physically bind to and downregulate CD4 in T cells, ensuring viral proliferation.<sup>87,88</sup>

Co-immunoprecipitation and mutational analyses suggest that residues 402–419 of CD4,

located in the membrane-proximal region of the cytoplasmic tail, are necessary for HIV-1 viral protein-induced downregulation.<sup>89–92</sup> Moreover, a putative amphipathic  $\alpha$ -helix in that region was found to be responsible for binding of both Nef and Vpu to CD4.<sup>91,93–96</sup> The CD4 receptor peptide (residues 403–419) was characterised by CD and NMR spectroscopy, and it was reported that residues 403–412 formed an  $\alpha$ -helix, with an equilibrium population of about 25%.<sup>97</sup> In a recent simulation study, employing REMD and MSM building techniques, it was demonstrated that the free energy landscape (FEL) for the receptor peptide (residues 402–419) was mainly flat—a characteristic feature of IDPs.<sup>98</sup>

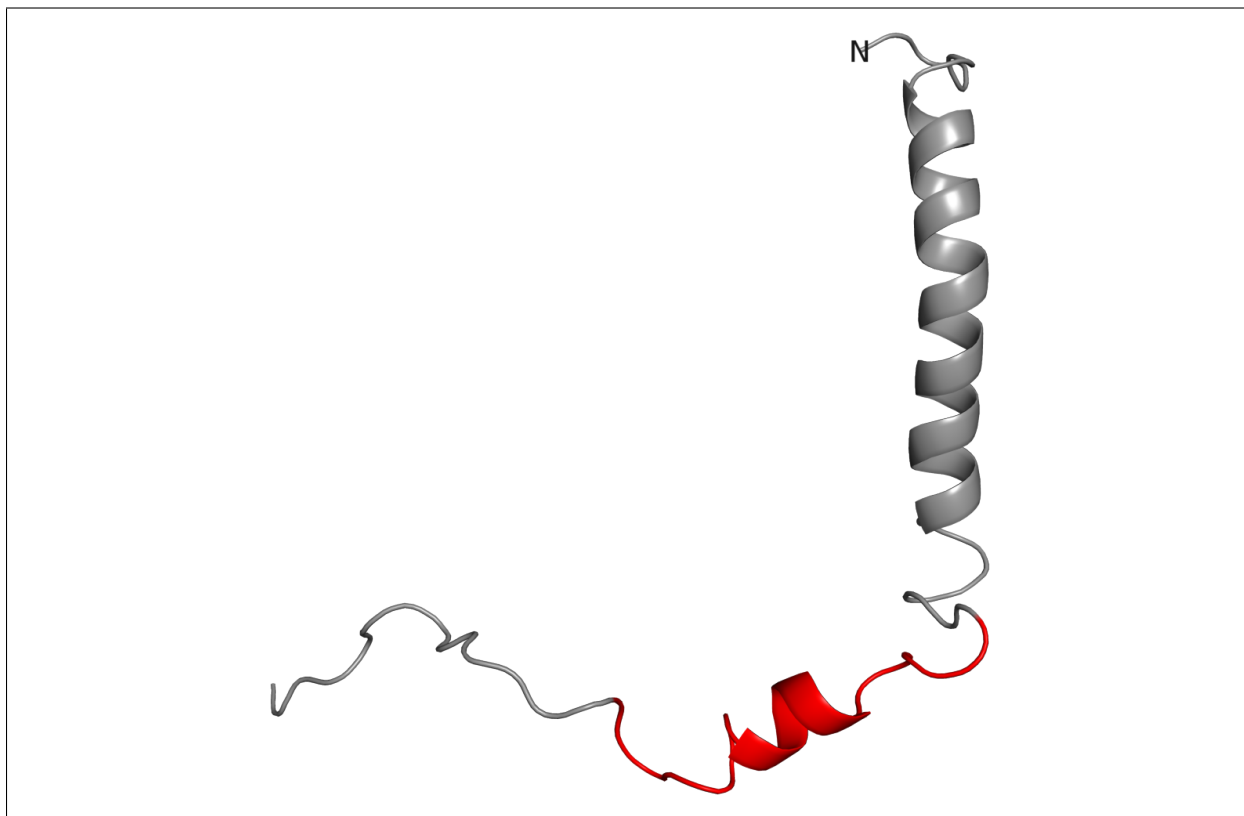


Figure 1: NMR solution structure for the transmembrane and cytoplasmic domains of human CD4 (PDB code: 2KLU).<sup>72</sup> The CD4 receptor peptide (402–419) is highlighted in red.

Here, an alternative approach for modelling intrinsic disorder in proteins, based on the potential energy landscape (PEL) framework,<sup>99–101</sup> is presented. Specifically, geometry optimisation-based approaches, namely basin-hopping parallel tempering (BHPT)<sup>102</sup> and

discrete path sampling (DPS),<sup>103–105</sup> are utilised to explore the PEL and FEL for the CD4 receptor peptide (CD4<sub>RP</sub>). Recently, Chebaro *et al.*<sup>106</sup> used REMD and DPS to probe the FEL for an IDR of the p53 upregulated modulator of apoptosis (PUMA) protein, and found that the potential energy landscape was inherently multifunnelled. In other work, BHPT was employed to predict favourable oligomers of amyloid  $\beta$ -peptide, A $\beta_{1-42}$ ,<sup>102</sup> implicated in Alzheimer’s disease.

The BHPT–DPS approach combines structure prediction and network building for efficient sampling of IDPs. Using these tools, benchmarks are presented for CD4<sub>RP</sub> ensembles generated with various state-of-the-art AMBER force fields (ff99SB-ILDN,<sup>45</sup> ff14ipq,<sup>107</sup> ff14SB<sup>108</sup>). AMBER force fields are widely used to model protein folding and there is an ongoing need to examine their performance, particularly for IDPs. Since the PEL framework offers an integrated approach for probing structure, dynamics and thermodynamics, along with powerful landscape visualisation techniques,<sup>99</sup> it can provide systematic comparisons of IDP ensembles generated by different force fields. We found that ff99SB-ILDN achieves a better balance of helical and random-coil structure for CD4<sub>RP</sub> than the more recent ff14ipq and ff14SB parametrisations. The free energy landscape for CD4<sub>RP</sub> was computed for ff99SB-ILDN, within the energy landscape framework, and various metastable states were identified. These results unify previous conflicting experimental findings for CD4<sub>RP</sub> and account for the rich functional repertoire of this intrinsically disordered region. Finally, we discuss the biological implications of our results relating to HIV-1 infection and opportunities for rational drug design.

## Methods

### Preparation of initial structures

The NMR solution structures for human CD4 (403–419) were obtained from the Protein Data Bank (PDB ID: 1WBR). All 32 models were extracted, and the missing arginine residue (402)

was added using PyMOL.<sup>109</sup> The N- and C-terminal ends of the peptide were capped with an acetyl group and methylamide, respectively.

### Force fields and solvent models

The performance of three AMBER force fields was tested in this study. AMBER ff99SB-ILDN<sup>45</sup> was developed to improve the side-chain torsion potentials of isoleucine, leucine, aspartate and asparagine in AMBER ff99SB.<sup>42</sup> Improvements of side-chain torsion potentials for most amino acids, as well as backbone  $\phi$  and  $\psi$  dihedral parameters, were later implemented as AMBER ff14SB.<sup>108</sup> Hence, these two force fields derive from the same parent (AMBER ff94).<sup>110</sup> In contrast, AMBER ff14ipq<sup>107</sup> represents a full rederivation of torsion parameters, coupled with the implicitly polarised charge (IPolQ) model<sup>111</sup> for approximating protein partial charges in the condensed phase.

Simulations of CD4<sub>RP</sub> (402–419) using the above force fields were performed with both explicit and implicit solvent models. For the explicit solvent investigations, ff99SB-ILDN and ff14SB force fields were paired with TIP3P,<sup>112</sup> and ff14ipq was paired with TIP4P-Ew.<sup>113</sup> The explicit solvent models were chosen to be consistent with the models used to parametrise the respective force fields.

All implicit solvent simulations were conducted by coupling the force fields with GB-Neck2<sup>114</sup> (igb=8 in Sander). Maffucci and Contini<sup>51</sup> reported that GB-Neck2 was able to compensate for slight helical biases displayed by the ff99SB series. The ff99SB-ILDN/GB-Neck2 combination was also able to discriminate helices from IDPs and reasonably predicted  $\beta$ -hairpins in their study.<sup>51</sup> An effective salt concentration of 0.1 M was maintained to provide mobile counterions in solution, and a cutoff of 25 Å was used for computation of the effective Born radius. The AMBER force fields were also correctly symmetrised,<sup>115</sup> to ensure that the energies of permutational isomers were identical.

## Structural refinement via MD

For each force field, the 32 initial structures were first minimised in vacuum for 8000 steepest-descent (SD) steps, followed by 2000 steps using the conjugate gradient (CG) method. The minimised structures were then solvated in an octahedral box, with a minimum distance of 12 Å between the peptide and the box edge. For each system, four Cl<sup>-</sup> ions were added to neutralise the charges on the peptide. The solvated structures were then minimised for 10000 steps (8000 SD; 2000 CG) with a cutoff of 10 Å and periodic boundary conditions for computing non-bonded interactions. A force constant of 100 kcal mol<sup>-1</sup> Å<sup>-2</sup> was used to restrain the peptide. The restraint was subsequently removed and the systems were allowed to relax for a further 10000 steps, using the same minimisation protocol as before.

Each system was then heated to 300 K for 20 ps, and the temperature was regulated using a Langevin thermostat with a collision frequency of 1 ps. A small restraint of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> was imposed on the peptide, and bonds involving hydrogen were constrained using the SHAKE algorithm, permitting an integration time step of 2 fs. The peptide restraints were removed and 5 ns of constant pressure (1 atm) MD was carried out; followed by 2 ns of MD in the canonical ensemble. Finally, 100 ns of MD in the NVT ensemble at 300 K was performed for each system (i.e. an aggregate production time of 3.2 μs for each force field). Frames along the trajectories were saved every 10 ps. The lowest energy conformer from each MD simulation (32 conformers per force field) was then selected to initiate structure prediction runs via basin-hopping parallel-tempering (BHPT).

## Structure prediction via BHPT

In BHPT, multiple basin-hopping (BH)<sup>116,117</sup> runs for replicas at different temperatures are performed simultaneously.<sup>102</sup> As with other replica exchange methods, the lower temperature limit for BHPT is usually chosen as the temperature at which physical observables are to be evaluated, while the high temperature limit is selected to permit crossing of the highest energy barriers on the landscape. After a given number of BH steps, replicas at adjacent

temperatures may be exchanged, based on a Metropolis criterion for the energies of the replicas.

In this work, the goal is not to locate the global potential energy minimum, since this structure for IDPs is unlikely to dominate the equilibrium properties. Instead, BHPT was utilised to explore the low energy regions of the potential energy landscape (PEL). The exchange of replicas in the BHPT procedure avoids kinetic trapping and facilitates accelerated exploration and sampling of the conformational space.

For each CD4<sub>RP</sub> conformer, BHPT runs were conducted using 16 replicas at temperatures exponentially distributed between 300 and 550 K using an implicit solvent representation. Each BHPT run consisted of 5000 BH steps and exchanges among neighbouring replicas were attempted every 10 steps. To explore the conformational space, random displacements of Cartesian coordinates were used (with a maximum step size of 1 Å) and step sizes were adjusted to achieve an acceptance probability of 22%. The most stable conformer at 300 K from each BHPT run was chosen as a starting point for construction of the PEL via discrete path sampling (DPS).<sup>103–105</sup>

### **Construction of potential energy landscapes via DPS**

Initial discrete paths between pairs of CD4<sub>RP</sub> conformers were first constructed. The starting conformer pairs (endpoints) were aligned using the LPERMDIST procedure in OPTIM.<sup>118</sup> This procedure effectively computes the minimised distance between endpoints with respect to translation, rotation and permutation. The updated quasi-continuous interpolation (QCI) scheme<sup>119</sup> was then used to predict configurations between the two endpoints. The interpolated geometries were then fed to the doubly-nudged<sup>120</sup> elastic band<sup>121,122</sup> (DNEB) routine, producing a discrete set of images between the selected endpoints. Maxima along this profile were then taken as initial transition state guesses. These candidates were then tightly converged using the hybrid eigenvector-following (HEF) procedure.<sup>123,124</sup> Refined transition states were then connected to minima by following steepest-descent paths parallel and anti-

parallel to the unique downhill direction. Minima were converged using a modified limited-memory Broyden-Fletcher-Golffarb-Shanno (L-BFGS) algorithm<sup>125,126</sup> with a convergence condition of  $10^{-7}$  kcal mol<sup>-1</sup> Å<sup>-1</sup> for the root-mean-square gradient. The initial paths were then optimised using the SHORTCUT<sup>127-129</sup> and UNTRAP<sup>128</sup> schemes in PATHSAMPLE<sup>130</sup> to produce the final kinetic transition networks (KTNs).

### Estimation of free energies and conversion pathways

The free energy landscape (FEL) was computed at 300 K using the harmonic superposition approximation (HSA).<sup>131</sup> Minima and transition states were first clustered into free energy groups using a recursive lumping procedure (REGROUPFREE in PATHSAMPLE).<sup>132</sup> The conversion pathways between conformers were obtained by applying Dijkstra’s shortest path algorithm<sup>127</sup> to the clustered KTN.

### Secondary structure analysis, computation of NMR shifts and coupling constants

Secondary structure analysis of minima in each transition network was conducted using the DSSP program<sup>133</sup> and NMR shifts were computed with ShiftX.<sup>134</sup> To determine overall NMR shifts, weighted sums, based on the equilibrium occupation probabilities, were calculated at 300 K for minima in the KTN.

Three-bond  $J_{HNH_\alpha}$  coupling constants were computed using the Karplus equation:

$$^3J_{HNH_\alpha}(\phi) = A\cos^2(\phi) + B\cos(\phi) + C, \tag{1}$$

$A = 9.5$ ,  $B = -1.4$  and  $C = 0.3$ , as proposed by Brüschweiler and Case.<sup>135</sup> Finally, the potential energy landscapes (and free energy surfaces) were visualised using disconnectivity graphs.<sup>136,137</sup>

# Results and Discussion

## Low-lying conformers

The NMR solution structures for residues 403–419 of CD4 are depicted in Figure 2a. The NMR-averaged ensemble displays notable structural variation, particularly in the C-terminal region, and the conformers show some tendency to form  $\alpha$ -helices between residues 403–412, as reported earlier by Willbold and Rösch.<sup>97</sup> Each of the 32 structures in the NMR ensemble were simulated via MD and subsequently BHPT using the three AMBER force fields. Explicit solvent MD was used to probe regions of local stability on the potential energy landscape of CD4<sub>RP</sub>. Seeding BHPT from MD trajectories is not strictly required, since BHPT runs can be initiated from the prepared NMR structures themselves. However, starting from locally stable regions on the PEL meant that fewer BHPT steps were required to satisfactorily explore low-lying wells on the PEL. On average, the potential energy of conformers improved by about 8 kcal mol<sup>-1</sup> after BHPT, and the final 32 conformers covered a range of less than 20 kcal mol<sup>-1</sup> for each of the three force fields.

The BHPT predicted conformers from each force field are shown in Figure 2. It is evident that ff99SB-ILDN and ff14ipq show greater structural variation than ff14SB. The conformers for the first two force fields resemble molten globules, in which there is partial formation of helical structure. Conversely, the final ensemble for ff14SB reveals significant helical character, where the CD4<sub>RP</sub> forms linear helices. On average, contiguous  $\alpha$ -helices nine, seven, and eleven residues long were obtained for ff99SB-ILDN, ff14ipq, ff14SB respectively. Notably, consistent with the NMR experiment, no antiparallel  $\beta$ -strands were seen.



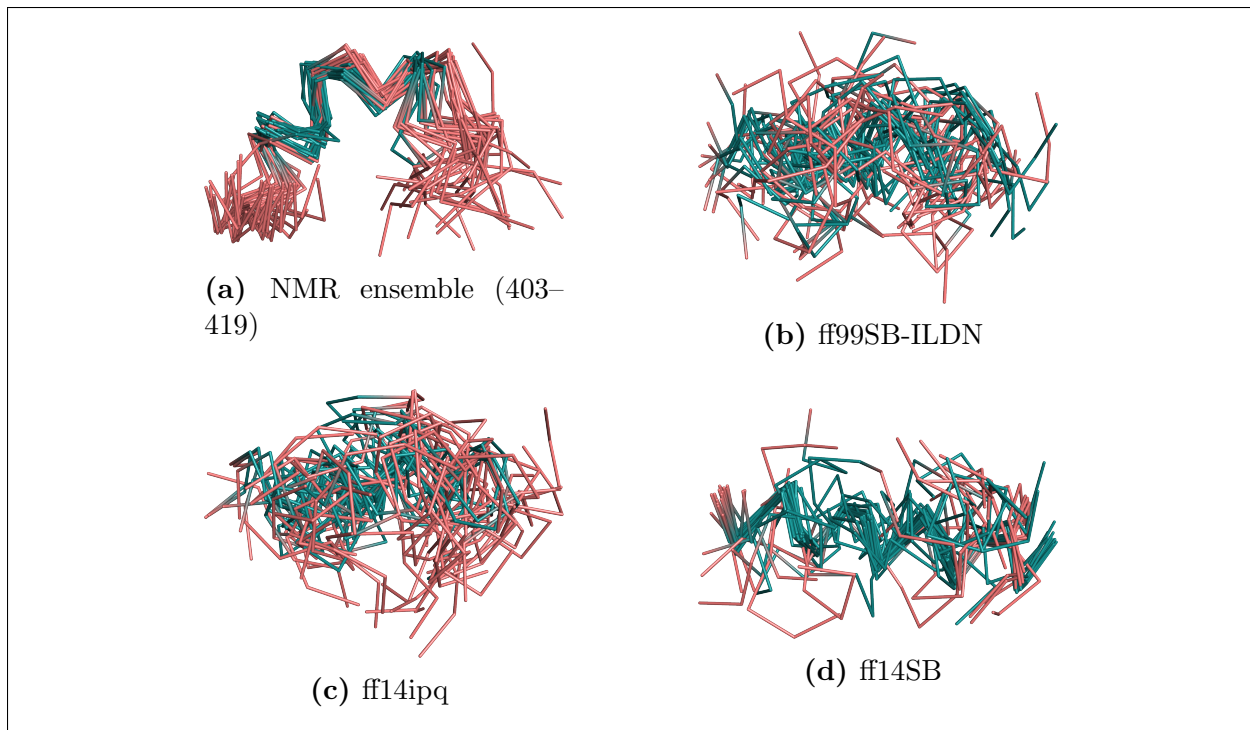
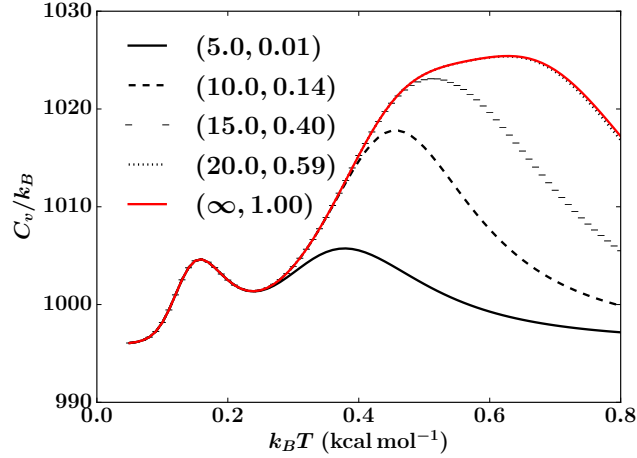


Figure 2: Structural ensembles for human CD4 receptor peptide: (a) 32 NMR solution structures (residues 403–419); (b)–(d) conformers (residues 402–419) obtained after BHPT, using various AMBER force fields. The conformers were clustered using PyMOL and were oriented with the N-terminal region on the left. The colouring scheme is: teal (helices), salmon (loops).

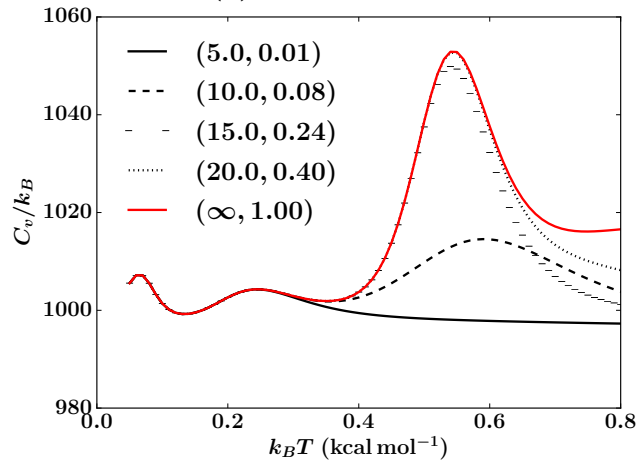
## Convergence of stationary point databases

The conformers in Figure 2b–2d were used to construct potential energy landscapes for  $\text{CD4}_{\text{RP}}$  via discrete path sampling. To test for convergence of the respective stationary point databases, the constant volume heat capacity ( $C_v$ ) curves were computed for subsets of the local minima. In particular,  $C_v$  curves were derived for minima within a given threshold above the global potential energy minimum (Figure 3). At low temperatures, we expect that features in the heat capacity may be attributed to a small subset of local minima. Hence, this approach should be robust in testing for convergence of the databases and should be able to reveal any deficiencies in sampling. For the three force fields tested, the low-temperature features in the heat capacity are converged when about 50% of all minima in

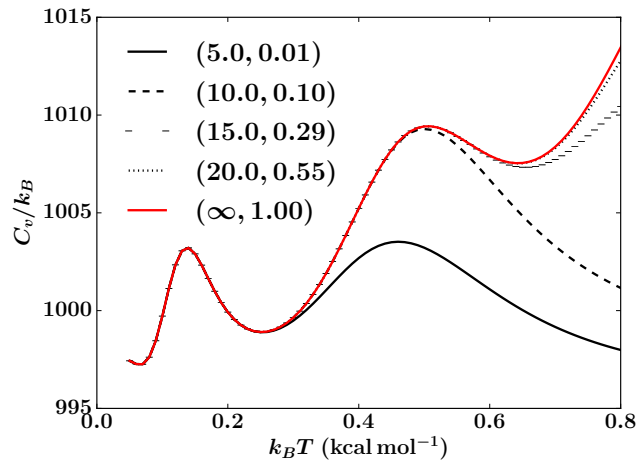
the corresponding stationary point database are included. A summary of the number of local minima and transition states in the final database for each force field is given in Table 1. In the next section, the properties of the resulting potential energy landscapes for CD4<sub>RP</sub> are compared systematically.



(a) ff99SB-ILDN



(b) ff14ipq



(c) ff14SB

Figure 3: Constant volume heat capacity curves ( $C_v$ ) for human CD4 receptor peptide. The curves were computed using the harmonic superposition approximation for the peptide simulated by various AMBER force fields. Minima within a specified energy threshold of the global potential energy minimum were incrementally included in the calculation of  $C_v$ . The energy thresholds ( $\text{kcal mol}^{-1}$ ) and the corresponding fraction of minima in the database are provided for each plot in parentheses.

**Table 1: Number of minima (MIN) and transition states (TS) in the DPS databases for the CD4 receptor peptide modelled by various AMBER force fields**

AMBER force field	No. of MIN	No. of TS
99SB-ILDN <sup>45</sup>	47434	51675
14ipq <sup>107</sup>	53926	58008
14SB <sup>108</sup>	67276	68689

## Characterisation of potential energy landscapes

The potential energy landscape for CD4<sub>RP</sub> modelled by the ff99SB-ILDN force field is presented in Figure 4; it is distinctively multifunnelled, with several prominent low-lying funnels. It was previously postulated that such multifunnelled potential energy landscapes are characteristic features of intrinsically disordered proteins.<sup>106</sup>

The disconnectivity graphs in Figure 4 are coloured based on the secondary structure content; specifically, the percent  $\alpha$ -helical, random-coil, and turn structure are shown, since these features were most common in the computed KTNs. The fractional  $\alpha$ -helicity is best able to classify the funnels on the landscape (Figure 4a). Many low energy conformers contain linear  $\alpha$ -helices, differing mainly in the length of the helical segment and orientation of side-chain groups. The shortest  $\alpha$ -helices in that region are about six residues long (approx. 1.7 turns), and the longest consists of 18 residues. Slightly higher in energy are conformers that display helical propensities in the N- and C-terminal segments of the peptide. There are also a number of conformers that lack  $\alpha$ -helical structure and are quite low in energy.

Chebaro *et al.* found that for the PUMA peptide the contiguous  $\alpha$ -helix was enthalpically unstable and that low-lying minima were relatively unstructured.<sup>106</sup> The organisation of the PEL for PUMA strongly suggested that the interaction of this IDP with molecular partners would mainly involve an induced-fit type mechanism. As mentioned above, the longest contiguous  $\alpha$ -helix (residues 402–419) in CD4<sub>RP</sub> corresponds to a low energy conformer on the PEL. This result suggests that interactions involving direct binding to CD4<sub>RP</sub> are most

likely to occur via conformational selection.

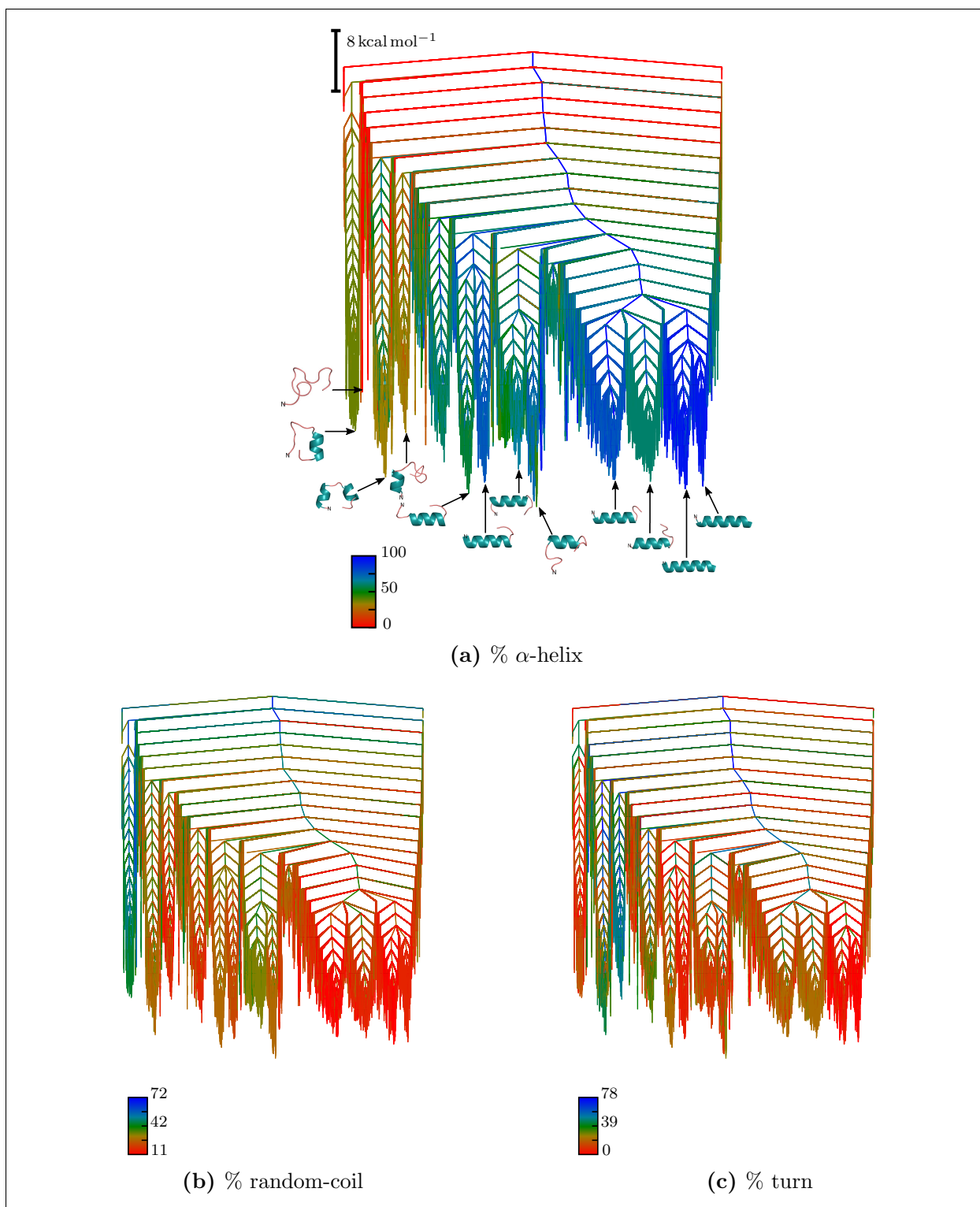


Figure 4: Potential energy landscape for CD<sub>4</sub><sub>RP</sub> modelled by the AMBER 99SB-ILDN force field. In the disconnectivity graphs, minima (branches) are coloured based on secondary structure content. Selected structures are superimposed on the graph and the  $\alpha$ -helical segments are coloured in teal.

The potential energy barriers between conformers exceeding 50%  $\alpha$ -helical content are much smaller than those separating conformers with residual helical content. The latter conformers generally lie higher in energy and display greater coil-like and turn secondary structure. Very few  $\beta$ -strand conformers were identified for the ff99SB-ILDN force field. For this representation, the conformers generally show  $\alpha$ -helical propensities, and only a few of them are largely unstructured.

Figure 5 illustrates the potential energy landscape for CD4<sub>RP</sub> computed with the first-generation ff14ipq force field. As for ff99SB-ILDN, the landscape is intrinsically disordered, with multiple prominent funnels. However, the relative potential energy barriers between the various low-lying conformers are notably larger than those obtained with ff99SB-ILDN. The conformers for ff14ipq are also more heterogeneous than those located with ff99SB-ILDN. The predicted linear  $\alpha$ -helical motifs in CD4<sub>RP</sub> are also shorter, by about two residues, for this force field. As before, conformers lacking significant helical structure, characterised by turns (Figure 5c) and coil-like (Figure 5b) secondary structure, generally reside in the high energy regions of the landscape. Additionally, a greater proportion of minima contain random-coil character than observed for ff99SB-ILDN.

Finally, the landscape for CD4<sub>RP</sub> was probed using the ff14SB force field; the topology is evidently less multifunnelled than for the former two force fields (Figure 6). Additionally, the landscape is dominated by linear  $\alpha$ -helical conformers. Secondary structure characterisation also reveals regions with substantial coil-like structure (Figure 6b), separated from the main part of the landscape by particularly high potential energy barriers. However, even in these regions, conformers still exhibit some helical character. Similar regions were also identified on the ff99SB-ILDN and ff14ipq landscapes, and the entropic barrier between them and the other parts of the PEL is therefore likely to be physically realistic.

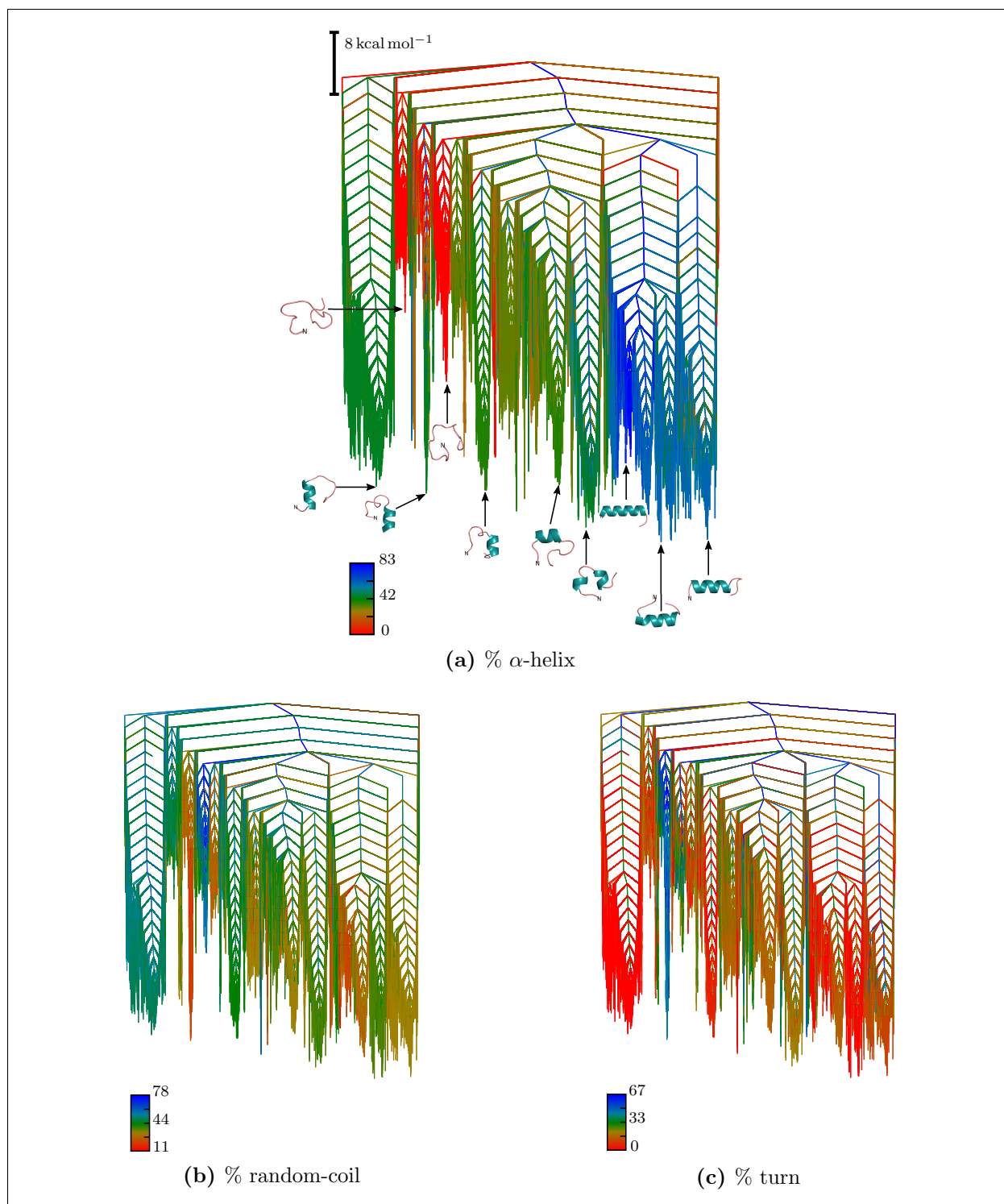


Figure 5: Potential energy landscape for CD4<sub>RP</sub> modelled by the AMBER 14ipq force field. In the disconnectivity graphs, minima (branches) are coloured based on secondary structure content. Selected structures are superimposed on the graph in (a) and the  $\alpha$ -helical segments are coloured in teal.



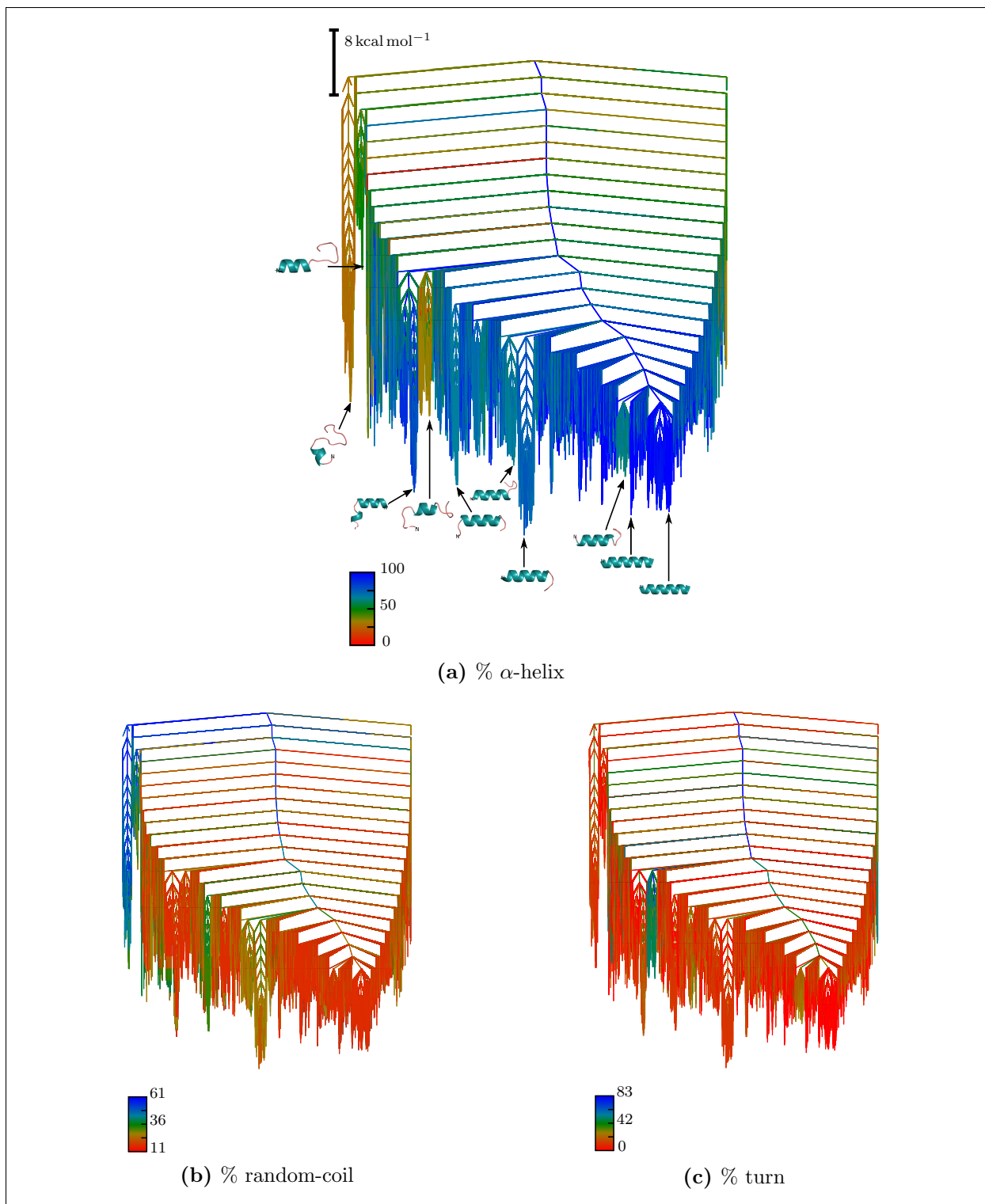


Figure 6: Potential energy landscape for CD<sub>4</sub><sub>RP</sub> modelled by the AMBER 14SB force field. In the disconnectivity graphs, minima (branches) are coloured based on secondary structure content. Selected structures are superimposed on the graph in (a) and the  $\alpha$ -helical segments are coloured in teal.

The distribution of the radius of gyration ( $R_g$ ) for the local minima of CD4<sub>RP</sub> is shown in Figure 7a. For all the force fields, there is a single peak in the  $R_g$  distribution. Conformers generated with ff14SB are most extended (i.e., corresponding to linear helices), while those predicted with ff14ipq are most compact (consistent with a greater presence of molten-globule like structures on the landscape). A slightly broader  $R_g$  distribution is obtained for ff99SB-ILDN, which peaks between the two newer AMBER force fields.

Figure 7b summarises the distribution of fractional  $\alpha$ -helicity. All force fields show multiple peaks in the distribution (at a bin width of 10%), with the major peaks for ff99SB-ILDN and ff14SB at about 60–70%  $\alpha$ -helicity. The distribution for ff14ipq is shifted to the left; in general, ff14ipq predicts peptides with fewer helical residues than the other two force fields. Additionally, on inspecting the left tails of the distributions, it is apparent that only a few conformers located with f14SB lack  $\alpha$ -helical character. The distributions for ff99SB-ILDN and ff14SB are somewhat similar, although the former parametrisation predicts a greater fraction of structures with low  $\alpha$ -helical content.

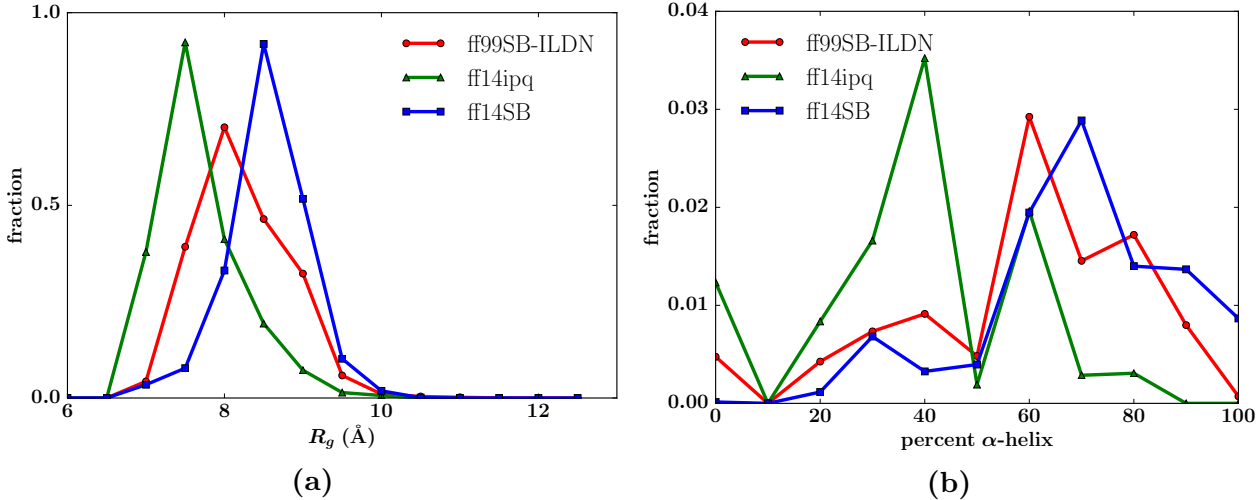


Figure 7: Distribution of radius of gyration ( $R_g$ ) and fractional  $\alpha$ -helicity for local minima of CD4<sub>RP</sub> generated with various AMBER force fields, as indicated.

## NMR shifts and coupling constants

Predicted NMR shifts are a useful tool for validating simulation data by comparison with experiment. The computed HN NMR shifts for human CD4 (residues 403–419) are presented in Figure 8. The agreement between the experimental and calculated NMR shifts was assessed by computing

$$\chi^2_\sigma = \frac{1}{N} \sum_{i=1}^N \frac{(\delta_{i,cal} - \delta_{i,exp})^2}{\sigma_\delta^2}, \quad (2)$$

where  $i$  is the residue number,  $N$  is the number of residues,  $\delta_{i,cal}$  and  $\delta_{i,exp}$  are the calculated (as described earlier) and experimental chemical shifts, respectively, and  $\sigma_\delta^2$  is the uncertainty ( $\sigma_\delta = 0.49$  ppm for HN using SHIFTX). By definition,  $\chi^2_\sigma > 1$  indicates a significant difference between computed and experimental structures.<sup>138</sup> The  $\chi^2_\sigma$  values for the HN NMR shifts for ff99SB-ILDN, ff14ipq and ff14SB structures are 0.28, 0.55 and 0.42, respectively. Based on this metric, the HN shifts for the ensembles generated with all three force fields agree quite well with the experimental findings.<sup>97</sup> The main deviations from the NMR experiments occur in the C-terminal region of the peptide (residues 417–419); the *de novo* approach employed in this work generally yields more ordered C-termini than found in experiment. In the N-terminal part of the peptide, ff14SB predicted more structured helices, while ff14ipq predicts more disorder than experiment. However, the overall HN NMR shifts, particularly those computed for ff99SB-ILDN, closely match the literature values.

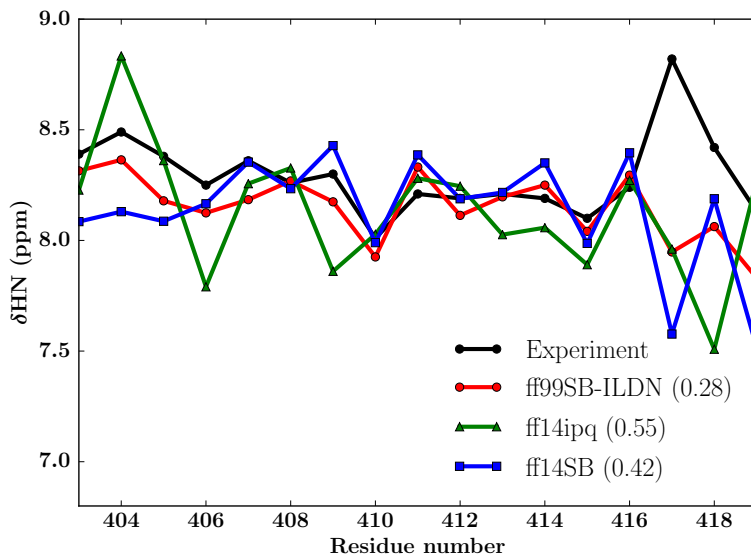


Figure 8: Comparison of HN NMR shift for  $\text{CD4}_{\text{RP}}$  ensembles generated with various AMBER force fields and experiment. The  $\chi^2_\sigma$  uncertainty, which measures the agreement between computed and experimental shifts, is given in parentheses.

In addition to NMR shifts, average  $^3J_{\text{HNH}_\alpha}$  values (weighted by occupation probabilities at 300 K) were computed for residues 403–419, and the root-mean-square deviation (rmsd) of the calculated and experimental values was determined. For ff99SB-ILDN, ff14ipq and ff14SB rms deviations of 1.40, 1.68 and 1.50 Hz were obtained, respectively. Overall, ff99SB-ILDN performs best for  $\text{CD4}_{\text{RP}}$ , in terms of reproducing the HN NMR shifts and  $J$  coupling constants. This force field was therefore selected for further analysis, specifically for probing the free energy surface.

## Free energy landscape for $\text{CD4}_{\text{RP}}$

The free energy landscape for  $\text{CD4}_{\text{RP}}$  was produced by clustering minima and transition states in the ff99SB-ILDN-derived kinetic transition network, based on the relative free energy barriers. To cluster the stationary points, the lowest energy minima with highest fractional random-coil and  $\alpha$ -helical structure were taken as the starting configurations for  $A$  (reactant) and  $B$  (product), respectively. Using a recursive regrouping scheme,  $A$  and  $B$  were then expanded as free energy groups, based on a predefined free energy threshold

( $\Delta F$ ) at 300 K. The rate constant for the  $A \rightarrow B$  conversion was derived by employing a graph transformation procedure.<sup>139</sup> To determine an appropriate value for  $\Delta F$ , regrouping was repeated for different values in the range 0 to 20 kcal mol<sup>-1</sup>, and the smallest threshold (in this case, 5.5 kcal mol<sup>-1</sup>) for which the rate constant converged was chosen for probing the free energy surface.

The resulting free energy landscape is depicted as a disconnectivity graph in Figure 9, and representative structures for various free energy groups are superimposed on the graph. The computed FEL is similar to the corresponding PEL in Figure 4. The lowest free energy group at 300 K (g1) contains a contiguous  $\alpha$ -helix extending from residues ALA404–SER415. Other free energy groups, of comparable energies, also contain linear  $\alpha$ -helical motifs: g2 (406–415), g4 (403–419), g5 (402–415).

A putative amphipathic  $\alpha$ -helix between residues 402–419 of the cytoplasmic tail of human CD4 may be necessary for interaction with both Nef and Vpu—HIV-1 accessory proteins.<sup>91,93–96</sup> Willbold and Rösch predicted an  $\alpha$ -helix from residues 403–412 for CD4<sub>RP</sub>.<sup>97</sup> Putative  $\alpha$ -helices extending from residues 402–417,<sup>69</sup> 406–415<sup>140</sup> and 404–413<sup>71,72</sup> have also been reported for the cytoplasmic tail of human CD4. On the computed free energy surface, these motifs are most similar to those found in g4, g2 and g9, respectively, which are all low-lying states. These results suggest that CD4<sub>RP</sub> is capable of adopting a wide range of helical motifs of varying lengths, which extend from different residues; however, the precise  $\alpha$ -helical structure observed is most likely dependent on the prevailing conditions or the resolution of the experiment. The results here help to unify previous experimental work, and explain the ability of CD4<sub>RP</sub> to bind different molecular partners. Additionally, Wittlich and colleagues reported that the amphipathic helix was stable even at 45°C.<sup>72</sup> The free energy landscape was also examined at 318 K (not shown) and is in agreement with those findings.

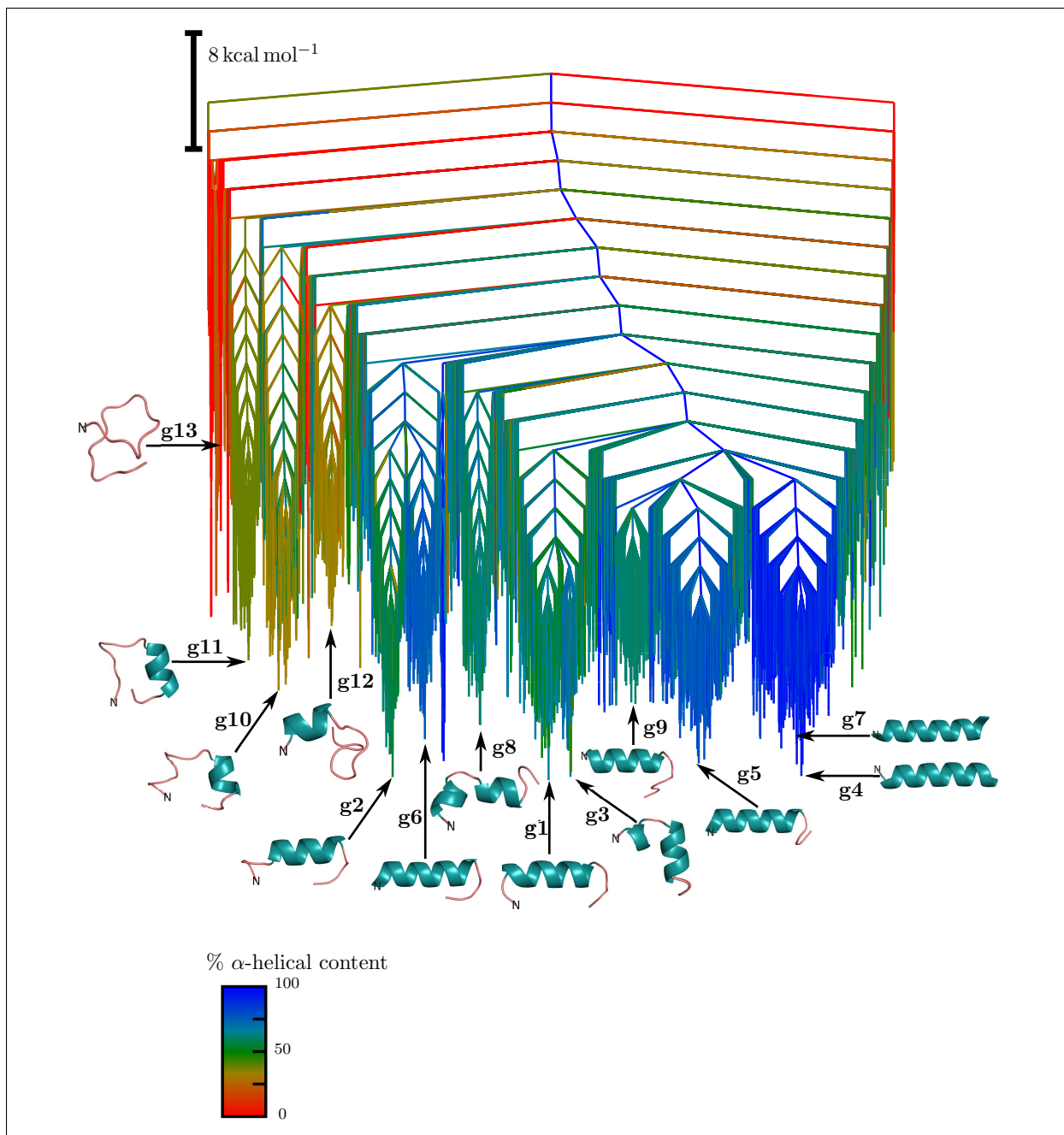


Figure 9: Free energy disconnectivity graph for CD4<sub>RP</sub> modelled by the ff99SB-ILDN force field. The free energy groups (branches) are coloured according to the  $\alpha$ -helical content. Representative structures of selected free energy groups are superimposed on the graph and the  $\alpha$ -helical motifs are coloured in teal.

The relatively high free energy barriers between the various  $\alpha$ -helical states on the FEL also suggest that conversion between them is likely to occur on long time scales. Furthermore,

fully unstructured states are generally higher in energy on the FEL and it is unlikely that CD4<sub>RP</sub> is unstructured under physiological conditions. Moreover, the intrinsic disorder in this peptide may be accurately described as a tendency to adopt various ordered states encompassing  $\alpha$ -helical scaffolds. The dominance of these  $\alpha$ -helical scaffolds in the low-lying regions of the FEL suggests that interaction with molecular partners may occur via conformational selection.

### Folding mechanism for CD4<sub>RP</sub>

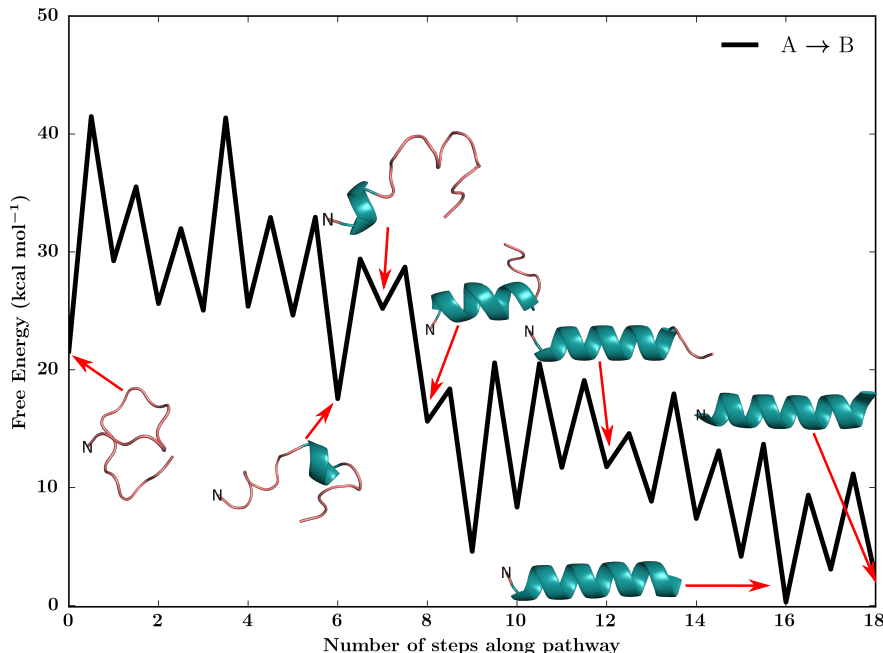


Figure 10: Folding pathway from a fully unstructured (*A*) state to a state containing a contiguous  $\alpha$ -helical structure 18 residues long (*B*) on the FEL for CD4<sub>RP</sub>. The number of steps corresponds to the number of transition states on the discrete path.

To gain further insight into the folding dynamics of CD4<sub>RP</sub> the pathway from a fully unstructured state (g13) to a state containing the longest contiguous  $\alpha$ -helix (g7; 18 residues long) was examined (Figure 10). Folding from the ‘random-coil-like’ state occurs in a cooperative fashion; i.e. there is a gradual decrease in the free energy of the states encountered as the peptide folds. This result agrees well with the findings of Ahalawat and co-workers,<sup>98</sup> who

also reported that the formation of the helix began at residues 407–410.<sup>98</sup> For the pathway depicted in Figure 10, the helix initiates from a similar region (410–413). The N-terminal portion folds first and several states on the pathway (between steps 8–12) exhibit  $\alpha$ -helical residues for the membrane proximal region (403–413) with an unstructured C-terminus. Interestingly, the helical motif in that segment of the path is almost identical in length and position with the original experimental prediction (403–412). It appears that this scaffold is an important structure for CD4<sub>RP</sub> function.

Finally, pathways and corresponding rate constants for conversions between various states on the FEL were examined. Generally, the rate for folding from states with linear helical motifs (e.g. g4, g9) to g1 is four to six orders of magnitude faster than from states without helices (e.g. g10, g11, g13). However, the rate constant for interconversions of various ordered helical scaffolds (e.g. g7  $\rightarrow$  g1) is of the order of  $10^{-5} \text{ s}^{-1}$ , suggesting that these transformations may occur slowly. The free energy barriers for such interconversions may be lowered in the presence of a molecular partner. Alternatively, different prevailing conditions in the cell may favour one state over the another, explaining why several groups have reported differing lengths for linear helical motifs for CD4<sub>RP</sub>.

# Conclusions

Binding of HIV-1 glycoprotein gp120 to the extracellular domain of CD4 leads to exposure of epitopes, which can be targeted by antibodies.<sup>141</sup> However, the presence of HIV-1 accessory proteins, Vpu and Nef, indirectly reduces the exposure of these epitopes by downregulating CD4 from the surface of infected cells.<sup>142,143</sup> Several approaches for HIV-1 antiviral drug design have focused on inducing exposure of HIV-1 epitopes, in the absence to CD4, to promote antibody-mediated responses.<sup>144–147</sup> These approaches adopt the conventional structure-based drug design model, which is based on targeting well-defined regions in proteins.<sup>148</sup>



Recently, some researchers suggested that an alternative route for drug development could be based on designing drugs that can mimic the interaction of the cytoplasmic tail of CD4 with HIV-1 accessory proteins, in an antagonistic fashion.<sup>147,149</sup> Similar design principles have already been adopted in cancer therapies, where small molecules that mimic intrinsically disordered regions have been used to inhibit critical protein-protein interactions.<sup>150,151</sup> However, the development of CD4 cytoplasmic tail-mimicking molecules has been slower, because the CD4-Vpu/Nef interaction site has been difficult to characterise.

In the present work, the free energy landscape for the CD4 receptor peptide has been explored by employing geometry optimisation-based approaches. The metastable states we have identified help to unify, and are consistent with, several earlier predictions. The conformers identified herein may be used as starting points in docking studies with HIV-1 accessory proteins, to better probe the recognition process. Additionally, the strategies presented in this work may be used to characterise the full-length cytoplasmic domain. These investigations may prove useful, not only in achieving thorough characterisation of the reaction interface, but may also represent critical step towards developing viable CD4 receptor peptide mimetics.

## Acknowledgement

The authors thank Brooke E. Husic and Debayn Chakraborty for useful comments and suggestions during the preparation of the manuscript. J. A. J acknowledges financial support from the Gates Cambridge Trust. D. J. W gratefully acknowledges support from the EPSRC.

## References

- (1) Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509–516.

- (2) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J. Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6661–6714.
- (3) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331.
- (4) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are natively unfolded proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 415–427.
- (5) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59.
- (6) Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533.
- (7) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. cell Biol.* **2005**, *6*, 197.
- (8) Dunker, A. K.; Romero, P.; Obradovic, Z.; Garner, E. C.; Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome Informatics* **2000**, *11*, 161–171.
- (9) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645.
- (10) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.

- (11) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins: Struct., Funct., Bioinf.* **2001**, *42*, 38–48.
- (12) Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. Flavors of protein disorder. *Proteins: Struct., Funct., Bioinf.* **2003**, *52*, 573–584.
- (13) Uversky, V. N.; Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804*, 1231–1264.
- (14) Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 932–951.
- (15) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **2002**, *41*, 6573–6582.
- (16) Demchenko, A. P. Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.* **2001**, *14*, 42–61.
- (17) Dyson, H. J.; Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
- (18) Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049.
- (19) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. cell Biol.* **2015**, *16*, 18.
- (20) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440.
- (21) Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **1996**, *35*, 13709–13715.

- (22) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584.
- (23) Baker, J. M. R.; Hudson, R. P.; Kanelis, V.; Choy, W.-Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738.
- (24) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5762–5767.
- (25) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246.
- (26) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* **2009**, *7*, e1000034.
- (27) Davey, N. E.; Travé, G.; Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **2011**, *36*, 159–169.
- (28) Dyson, H. J.; Wright, P. E. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* **2001**, *339*, 258–270.
- (29) Dyson, H. J.; Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **2004**, *104*, 3607–3622.
- (30) Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.

- (31) Kikhney, A. G.; Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577.
- (32) Ferreon, A. C. M.; Moran, C. R.; Gambin, Y.; Deniz, A. A. Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol.* **2010**, *472*, 179–204.
- (33) Miyagi, A.; Tsunaka, Y.; Uchihashi, T.; Mayanagi, K.; Hirose, S.; Morikawa, K.; Ando, T. Visualization of intrinsically disordered regions of proteins by highspeed atomic force microscopy. *ChemPhysChem* **2008**, *9*, 1859–1866.
- (34) Bodart, J.-F.; Wieruszeski, J.-M.; Amniai, L.; Leroy, A.; Landrieu, I.; Rousseau-Lescuyer, A.; Vilain, J.-P.; Lippens, G. NMR observation of Tau in *Xenopus* oocytes. *J. Magn. Reson.* **2008**, *192*, 252–257.
- (35) Iakoucheva, L. M.; Kimzey, A. L.; Masselon, C. D.; Bruce, J. E.; Garner, E. C.; Brown, C. J.; Dunker, A. K.; Smith, R. D.; Ackerman, E. J. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* **2001**, *10*, 560–571.
- (36) Rauscher, S.; Pomès, R. Molecular simulations of protein disorder. *Biochem. Cell Biol.* **2010**, *88*, 269–290.
- (37) Baker, C. M.; Best, R. B. Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 182–198.
- (38) Levine, Z. A.; Shea, J.-E. Simulations of disordered proteins and systems with conformational heterogeneity. *Curr. Opin. Struct. Biol.* **2017**, *43*, 95–103.
- (39) Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147–154.

- (40) Vendruscolo, M. Determination of conformationally heterogeneous states of proteins. *Curr. Opin. Struct. Biol.* **2007**, *17*, 15–20.
- (41) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116.
- (42) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (43) Nguyen, P. H.; Li, M. S.; Derreumaux, P. Effects of all-atom force fields on amyloid oligomerization: Replica exchange molecular dynamics simulations of the A $\beta$  1622 dimer and trimer. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9778–9788.
- (44) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (45) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E.; LindorffLarsen, K.; Piana, S.; Palmo, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.
- (46) Best, R. B.; Mittal, J. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.
- (47) Wang, W.; Ye, W.; Jiang, C.; Luo, R.; Chen, H. New force field on modeling intrinsically disordered proteins. *Chem. Biol. Drug Des.* **2014**, *84*, 253–269.
- (48) Palazzesi, F.; Prakash, M. K.; Bonomi, M.; Barducci, A. Accuracy of current all-atom force-fields in modeling protein disordered states. *J. Chem. Theory Comput.* **2015**, *11*, 2–7.

- (49) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.
- (50) Smith, M. D.; Rao, J. S.; Segelken, E.; Cruz, L. Force-field induced bias in the structure of A $\beta$ 2130: A comparison of OPLS, AMBER, CHARMM, and GROMOS force fields. *J. Chem. Inf. Model.* **2015**, *55*, 2587–2595.
- (51) Maffucci, I.; Contini, A. An updated test of AMBER force fields and implicit solvent models in predicting the secondary structure of helical,  $\beta$ -hairpin, and intrinsically disordered peptides. *J. Chem. Theory Comput.* **2016**, *12*, 714–727.
- (52) Fisher, C. K.; Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426–431.
- (53) Schor, M.; Mey, A. S. J. S.; MacPhee, C. E. Analytical methods for structural ensembles and dynamics of intrinsically disordered proteins. *Biophys. Rev.* **2016**, *8*, 429–439.
- (54) Chong, S.-H.; Chatterjee, P.; Ham, S. Computer Simulations of Intrinsically Disordered Proteins. *Annu. Rev. Phys. Chem.* **2017**, *68*, 117–134.
- (55) Ostermeir, K.; Zacharias, M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 847–853.
- (56) Michel, J.; Cuchillo, R. The impact of small molecule binding on the energy landscape of the intrinsically disordered protein C-myc. *PLoS One* **2012**, *7*, e41070.
- (57) Baftizadeh, F.; Pietrucci, F.; Biarnés, X.; Laio, A. Nucleation process of a fibril precursor in the C-terminal segment of amyloid- $\beta$ . *Phys. Rev. Lett.* **2013**, *110*, 168103.

- (58) Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **2015**, *5*, 15449.
- (59) Zerze, G. H.; Miller, C. M.; Granata, D.; Mittal, J. Free energy surface of an intrinsically disordered protein: comparison between temperature replica exchange molecular dynamics and bias-exchange metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2776–2782.
- (60) Lin, Y.-S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophys. J.* **2012**, *102*, 315–324.
- (61) Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation. *J. Am. Chem. Soc.* **2013**, *135*, 16092–16101.
- (62) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 07B604\_1.
- (63) Stanley, N.; Esteban-Martín, S.; De Fabritiis, G. Kinetic modulation of a disordered protein domain by phosphorylation. *Nat. Commun.* **2014**, *5*, 5272.
- (64) Garcia-Pino, A. Intrinsic disordered controls transcription of bacterial toxin-antitoxin modules. *Biophys. J.* **2017**, *112*, 317a–318a.
- (65) Choi, J.-M.; Wang, J.; Holehouse, A. S.; Alberti, S.; Hyman, A. A.; Pappu, R. V. Molecular grammar governing phase behavior of intrinsically disordered proteins with prion-like domains. *Biophys. J.* **2018**, *114*, 561a.



- (66) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (67) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- (68) Husic, B. E.; Pande, V. S. Markov State Models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.
- (69) Wray, V.; Mertins, D.; Kiess, M.; Henklein, P.; Trowitzsch-Kienast, W.; Schubert, U. Solution structure of the cytoplasmic domain of the human CD4 glycoprotein by CD and <sup>1</sup>H NMR spectroscopy: implications for biological functions. *Biochemistry* **1998**, *37*, 8527–8538.
- (70) Zhu, J.; Paul, W. E. CD4 T cells: fates, functions, and faults. *Blood* **2008**, *112*, 1557–1569.
- (71) Wittlich, M.; Koenig, B. W.; Hoffmann, S.; Willbold, D. Structural characterization of the transmembrane and cytoplasmic domains of human CD4. *Biochim. Biophys. Acta, Biomembr* **2007**, *1768*, 2949–2960.
- (72) Wittlich, M.; Thiagarajan, P.; Koenig, B. W.; Hartmann, R.; Willbold, D. NMR structure of the transmembrane and cytoplasmic domains of human CD4 in micelles. *Biochim. Biophys. Acta, Biomembr* **2010**, *1798*, 122–127.
- (73) Doyle, C.; Strominger, J. L. Interaction between CD4 and class II MHC molecules mediates cell adhesion. *Nature* **1987**, *330*, 256.
- (74) Sleckman, B. P.; Peterson, A.; Jones, W. K.; Foran, J. A.; Greenstein, J. L.; Seed, B.; Burakoff, S. J. Expression and function of CD4 in a murine T-cell hybridoma. *Nature* **1987**, *328*, 351.

- (75) Turner, J. M.; Brodsky, M. H.; Irving, B. A.; Levin, S. D.; Perlmutter, R. M.; Littman, D. R. Interaction of the unique N-terminal region of tyrosine kinase p56lck with cytoplasmic domains of CD4 and CD8 is mediated by cysteine motifs. *Cell* **1990**, *60*, 755–765.
- (76) Straus, D. B.; Weiss, A. Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor. *Cell* **1992**, *70*, 585–593.
- (77) Chan, A. C.; Iwashima, M.; Turck, C. W.; Weiss, A. ZAP-70: a 70 kd protein-tyrosine kinase that associates with the TCR  $\zeta$  chain. *Cell* **1992**, *71*, 649–662.
- (78) Dalgleish, A. G.; Beverley, P. C. L.; Clapham, P. R.; Crawford, D. H.; Greaves, M. F.; Weiss, R. A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **1984**, *312*, 763.
- (79) Klatzmann, D.; Champagne, E.; Chamaret, S.; Gruest, J.; Guetard, D.; Hercend, T.; Gluckman, J.-C.; Montagnier, L. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **1984**, *312*, 767.
- (80) Maddon, P. J.; Dalgleish, A. G.; McDougal, J. S.; Clapham, P. R.; Weiss, R. A.; Axel, R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **1986**, *47*, 333–348.
- (81) Sattentau, Q. J.; Dalgleish, A. G.; Weiss, R. A.; Beverley, P. C. Epitopes of the CD4 antigen and HIV infection. *Science* **1986**, *234*, 1120–1123.
- (82) McDougal, J. S.; Kennedy, M. S.; Sligh, J. M.; Cort, S. P.; Mawle, A.; Nicholson, J. K. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **1986**, *231*, 382–385.
- (83) Robinson, H. L.; Zinkus, D. M. Accumulation of human immunodeficiency virus type 1 DNA in T cells: results of multiple infection events. *J. Virol.* **1990**, *64*, 4836–4841.

- (84) Pauza, C. D.; Galindo, J. E.; Richman, D. D. Reinfection results in accumulation of unintegrated viral DNA in cytopathic and persistent human immunodeficiency virus type 1 infection of CEM cells. *J. Exp. Med.* **1990**, *172*, 1035–1042.
- (85) Ross, T. M.; Oran, A. E.; Cullen, B. R. Inhibition of HIV-1 progeny virion release by cell-surface CD4 is relieved by expression of the viral Nef protein. *Curr. Biol.* **1999**, *9*, 613–621.
- (86) Levesque, K.; Zhao, Y.-S.; Cohen, É. A. Vpu exerts a positive effect on HIV-1 infectivity by down-modulating CD4 receptor molecules at the surface of HIV-1-producing cells. *J. Biol. Chem.* **2003**, *278*, 28346–28353.
- (87) Garcia, J. V.; Miller, A. D. Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* **1991**, *350*, 508.
- (88) Willey, R. L.; Maldarelli, F.; Martin, M. A.; Strebel, K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. *J. Virol.* **1992**, *66*, 7193–7200.
- (89) Chen, M. Y.; Maldarelli, F.; Karczewski, M. K.; Willey, R. L.; Strebel, K. Human immunodeficiency virus type 1 Vpu protein induces degradation of CD4 in vitro: the cytoplasmic domain of CD4 contributes to Vpu sensitivity. *J. Virol.* **1993**, *67*, 3877–3884.
- (90) Vincent, M. J.; Raja, N. U.; Jabbar, M. A. Human immunodeficiency virus type 1 Vpu protein induces degradation of chimeric envelope glycoproteins bearing the cytoplasmic and anchor domains of CD4: role of the cytoplasmic domain in Vpu-induced degradation in the endoplasmic reticulum. *J. Virol.* **1993**, *67*, 5538–5549.
- (91) Aiken, C.; Konner, J.; Landau, N. R.; Lenburg, M. E.; Trono, D. Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **1994**, *76*, 853–864.

- (92) Anderson, S. J.; Lenburg, M.; Landau, N. R.; Garcia, J. V. The cytoplasmic domain of CD4 is sufficient for its down-regulation from the cell surface by human immunodeficiency virus type 1 Nef. *J. Virol.* **1994**, *68*, 3092–3101.
- (93) Yao, X.-J.; Friberg, J.; Checroune, F.; Gratton, S.; Boisvert, F.; Sékaly, R. P.; Cohen, E. A. Degradation of CD4 induced by human immunodeficiency virus type 1 Vpu protein: a predicted alpha-helix structure in the proximal cytoplasmic region of CD4 contributes to Vpu sensitivity. *Virology* **1995**, *209*, 615–623.
- (94) Tiganos, E.; Yao, X.-J.; Friberg, J.; Daniel, N.; Cohen, E. A. Putative alpha-helical structures in the human immunodeficiency virus type 1 Vpu protein and CD4 are involved in binding and degradation of the CD4 molecule. *J. Virol.* **1997**, *71*, 4452–4460.
- (95) Preusser, A.; Briese, L.; Willbold, D. Presence of a helix in human CD4 cytoplasmic domain promotes binding to HIV-1 Nef protein. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 734–740.
- (96) Singh, S. K.; Möckel, L.; Thiagarajan-Rosenkranz, P.; Wittlich, M.; Willbold, D.; Koenig, B. W. Mapping the interaction between the cytoplasmic domains of HIV1 viral protein U and human CD4 with NMR spectroscopy. *FEBS J.* **2012**, *279*, 3705–3714.
- (97) Willbold, D.; Rösch, P. Solution structure of the human CD4 (403–419) receptor peptide. *J. Biomed. Sci.* **1996**, *3*, 435–441.
- (98) Ahalawat, N.; Arora, S.; Murarka, R. K. Structural ensemble of CD4 cytoplasmic tail (402419) reveals a nearly flat free-energy landscape with local  $\alpha$ -helical order in aqueous solution. *J. Phys. Chem. B* **2015**, *119*, 11229–11242.
- (99) Wales, D. J. *Energy landscapes: Applications to clusters, biomolecules and glasses*; Cambridge University Press: Cambridge, U.K., 2003.

- (100) Wales, D. J. Energy landscapes and properties of biomolecules. *Phys. Biol.* **2005**, *2*, S86.
- (101) Wales, D. J. Energy landscapes: some new horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.
- (102) Strodel, B.; Lee, J. W. L.; Whittleston, C. S.; Wales, D. J. Transmembrane structures for Alzheimer’s A $\beta$ 1 42 oligomers. *J. Am. Chem. Soc.* **2010**, *132*, 13300–13312.
- (103) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3305.
- (104) Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.
- (105) Evans, D. A.; Wales, D. J. Folding of the GB1 hairpin peptide from discrete path sampling. *J. Chem. Phys.* **2004**, *121*, 1080–1090.
- (106) Chebaro, Y.; Ballard, A. J.; Chakraborty, D.; Wales, D. J. Intrinsically disordered energy landscapes. *Sci. Rep.* **2015**, *5*, 10386.
- (107) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. ff14ipq: a self-consistent force field for condensed-phase simulations of proteins. *J. Chem. Theory Comput.* **2014**, *10*, 4515–4534.
- (108) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (109) Schrödinger, L.
- (110) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

- (111) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **2013**, *117*, 2328–2338.
- (112) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (113) Horn, H. W.; Swope, W. C.; Pitara, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (114) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved generalized born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.
- (115) Małolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S. N.; Wales, D. J. Symmetrization of the AMBER and CHARMM force fields. *J. Comput. Chem.* **2010**, *31*, 1402–1409.
- (116) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611–6615.
- (117) Wales, D. J.; Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (118) Wales, D. J. OPTIM: A program for optimising geometries and calculating pathways. <http://www-wales.ch.cam.ac.uk/OPTIM>.
- (119) Röder, K.; Wales, D. J.; Roder, K.; Wales, D. J. Predicting pathways between distant configurations for biomolecules. *J. Chem. Theory Comput.* **2018**, *14*, 4271–4278.

- (120) Trygubenko, S. A.; Wales, D. J. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **2004**, *120*, 2082–2094.
- (121) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
- (122) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
- (123) Munro, L. J.; Wales, D. J. Defect migration in crystalline silicon. *Phys. Rev. B* **1999**, *59*, 3969.
- (124) Kumeda, Y.; Wales, D. J.; Munro, L. J. Transition states and rearrangement mechanisms from hybrid eigenvector-following and density functional theory.: application to C<sub>10</sub>H<sub>10</sub> and defect migration in crystalline silicon. *Chem. Phys. Lett.* **2001**, *341*, 185–194.
- (125) Nocedal, J. Updating quasi-Newton matrices with limited storage. *Math. Comput.* **1980**, *35*, 773–782.
- (126) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **1989**, *45*, 503–528.
- (127) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. Finding pathways between distant local minima. *J. Chem. Phys.* **2005**, *122*, 234903.
- (128) Strodel, B.; Whittleston, C. S.; Wales, D. J. Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.
- (129) Wales, D. J.; Carr, J. M.; Khalili, M.; de Souza, V. K.; Strodel, B.; Whittleston, C. S.

- Proteins Energy, Heat Signal Flow*; Computation in chemistry; CRC Press, 2009; p 315.
- (130) Wales, D. J. PATHSAMPLE: A driver for OPTIM to create stationary point databases using discrete path sampling and perform kinetic analysis. <http://www-wales.ch.cam.ac.uk/PATHSAMPLE>.
  - (131) Strodel, B.; Wales, D. J. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.* **2008**, *466*, 105–115.
  - (132) Whittleston, C. S. Energy landscapes of biological systems. Ph.D. thesis, University of Cambridge, 2012.
  - (133) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogenbonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
  - (134) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and accurate calculation of protein  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts. *J. Biomol. NMR* **2003**, *26*, 215–240.
  - (135) Brüschweiler, R.; Case, D. A. Adding harmonic motion to the Karplus relation for spin-spin coupling. *J. Am. Chem. Soc.* **1994**, *116*, 11199–11200.
  - (136) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517.
  - (137) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.
  - (138) Ball, K. A.; Wemmer, D. E.; Head-Gordon, T. Comparison of structure determination methods for intrinsically disordered amyloid- $\beta$  peptides. *J. Phys. Chem. B* **2014**, *118*, 6405–6416.



- (139) Wales, D. J. Calculating rate constants and committor probabilities for transition networks by graph transformation. *J. Chem. Phys.* **2009**, *130*, 204111.
- (140) Kim, P. W.; Sun, Z.-Y. J.; Blacklow, S. C.; Wagner, G.; Eck, M. J. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science* **2003**, *301*, 1725–1728.
- (141) Veillette, M.; Désormeaux, A.; Medjahed, H.; Gharsallah, N.-E.; Coutu, M.; Baalwa, J.; Guan, Y.; Lewis, G.; Ferrari, G.; Hahn, B. H. Interaction with cellular CD4 exposes HIV-1 envelope epitopes targeted by antibody-dependent cell-mediated cytotoxicity. *J. Virol.* **2014**, *88*, 2633–2644.
- (142) Arias, J. F.; Heyer, L. N.; von Bredow, B.; Weisgrau, K. L.; Moldt, B.; Burton, D. R.; Rakasz, E. G.; Evans, D. T. Tetherin antagonism by Vpu protects HIV-infected cells from antibody-dependent cell-mediated cytotoxicity. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6425–6430.
- (143) Veillette, M.; Coutu, M.; Richard, J.; Batrville, L.-A.; Dagher, O.; Bernard, N.; Tremblay, C.; Kaufmann, D. E.; Roger, M.; Finzi, A. The HIV-1 gp120 CD4-bound conformation is preferentially targeted by antibody-dependent cellular cytotoxicity-mediating antibodies in sera from HIV-1-infected individuals. *J. Virol.* **2015**, *89*, 545–551.
- (144) Martin, L.; Stricher, F.; Missé, D.; Sironi, F.; Pugnière, M.; Barthe, P.; Prado-Gotor, R.; Freulon, I.; Magne, X.; Roumestand, C. Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat. Biotechnol.* **2003**, *21*, 71.
- (145) Haim, H.; Si, Z.; Madani, N.; Wang, L.; Courter, J. R.; Princiotto, A.; Kassa, A.; DeGrace, M.; McGee-Estrada, K.; Mefford, M. Soluble CD4 and CD4-mimetic com-

- pounds inhibit HIV-1 infection by induction of a short-lived activated state. *PLoS Pathog.* **2009**, *5*, e1000360.
- (146) Baleux, F.; Loureiro-Morais, L.; Hersant, Y.; Clayette, P.; Arenzana-Seisdedos, F.; Bonnaffé, D.; Lortat-Jacob, H. A synthetic CD4heparan sulfate glycoconjugate inhibits CCR5 and CXCR4 HIV-1 attachment and entry. *Nat. Chem. Biol.* **2009**, *5*, 743.
- (147) Richard, J.; Veillette, M.; Brassard, N.; Iyer, S. S.; Roger, M.; Martin, L.; Pazgier, M.; Schön, A.; Freire, E.; Routy, J.-P. CD4 mimetics sensitize HIV-1-infected cells to ADCC. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2687–E2694.
- (148) Cheng, Y.; LeGall, T.; Oldfield, C. J.; Mueller, J. P.; Van, Y.-Y. J.; Romero, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Rational drug design via intrinsically disordered protein. *Trends Biotechnol.* **2006**, *24*, 435–442.
- (149) Curreli, F.; Do Kwon, Y.; Zhang, H.; Yang, Y.; Scacalossi, D.; Kwong, P. D.; Deb-nath, A. K. Binding mode characterization of NBD series CD4-mimetic HIV-1 entry inhibitors by X-ray structure and resistance study. *Antimicrob. Agents Chemother.* **2014**, *58*, 5478–5491.
- (150) Arkin, M. Protein-protein interactions and cancer: small molecules going in for the kill. *Curr. Opin. Chem. Biol.* **2005**, *9*, 317–324.
- (151) Fry, D. C.; Vassilev, L. T. Targeting protein-protein interactions for cancer therapy. *J. Mol. Med.* **2005**, *83*, 955–963.

## Graphical TOC Entry

